UNITED STATES PATENT APPLICATION

OF

WILLIAM G. HARLESS

MICHAEL G. HARLESS

and

MARCIA A. ZIER

FOR

INTERACTIVE SIMULATED DIALOGUE SYSTEM AND METHOD FOR A COMPUTER NETWORK

BACKGROUND OF THE INVENTION

The present invention relates generally to an interactive simulated dialogue system and method for simulating a dialogue between persons. More particularly, the present invention relates to an audiovisual simulated dialogue system and method for providing a simulated dialogue over a computer network. Currently, a simulated dialogue program combines digital video and voice recognition technology to allow a user to speak naturally and conduct a virtual interview with images of a human character. These programs facilitate, for example, professional education through direct virtual dialogue with acknowledged experts; patient education through direct virtual dialogue with health professionals and experienced peers; and foreign language training through virtual interviews with native speakers.

Simulated dialogue programs have been developed in accordance with the methods and apparatus disclosed by *Harless*, U.S. Patent No 5,006,987. One such program is a virtual interview with Dr. Jackie Johnson, a female oncologist, which allows women concerned about breast cancer to obtain in-depth information from this acknowledged expert. Another simulated dialogue program allows users to learn about the issues and concerns of biological warfare from Dr. Joshua Lederberg, a Nobel laureate. Still another program allows students of the Arabic language to conduct virtual interviews with Iraqi native speakers to learn conversational Arabic and sustain their proficiency with that language.

These programs, however, are implemented in a stand-alone computer environment. As such, each user must not only have the necessary hardware, they also need to install the necessary software. Moreover, the users must choose and select the desired simulation topics to be loaded on the computer as well as supplement them on an ongoing basis. Thus, it is desirable to provide realistic simulated dialogues over a computer network.

SUMMARY OF THE INVENTION

Accordingly, the present invention is directed to an interactive simulated dialogue system that substantially obviates one or more of the problems due to limitations and disadvantages of the related art.

In accordance with the purposes of the present invention, as embodied and broadly described, the invention provides a system for an interactive simulated dialogue over a network including a client node connected to the network including a browser for selecting a simulated dialogue program, a network connection for receiving over the network a vocabulary set corresponding to the selected simulation program, a client agent transmitting over the network signals corresponding to a user voice input, a client buffer agent receiving over the network signals representative of a meaningful response to the user voice input, and an output component for outputting an audiovisual representation of a human being speaking the meaningful response. The system further includes a server coupled to

the network including a database containing vocabulary sets, wherein each vocabulary set corresponds to a simulated dialogue program, a server launch agent receiving over the network the selected simulated dialogue program and transmitting over the network the vocabulary set corresponding to the selected simulated dialogue program, a server agent for receiving signals over the network corresponding to the user voice input and for determining a meaningful response to the user voice input, and a server buffer agent for transmitting over the network signals representative of the meaningful response.

In another embodiment, the invention provides a method for an interactive simulated dialogue over a computer network including a client node and a server. The method performed by the client node includes determining a system capacity of the client node, receiving a simulated dialogue program from the server, installing the simulated dialogue program based on the determination of the system capacity, receiving user voice input, transmitting to the server signals corresponding to the user voice input, receiving from the server signals representative of a meaningful response to the user voice input, and outputting an audiovisual representation of a human being speaking the meaningful response.

The accompanying drawings are included to provide a further understanding of the invention and are incorporated in and constitute a part of this specification, illustrate several embodiments of the invention and together with the description serve to explain the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated in and constitute a part of this specification, illustrate one embodiment of the invention and together with the description, serve to explain the principles of the invention.

In the drawings,

Fig. 1 is a schematic diagram of an interactive simulated dialogue system over a computer network according to one embodiment of the present invention;

Fig. 2 is a schematic diagram illustrating in detail the query process shown in Fig. 1;

Fig. 3 is a general flow diagram of the interactive simulation;

Fig. 4 is a detailed flow diagram of the client node;

Fig. 5 is a detailed flow diagram of the server; and

Fig. 6 shows a relationship between an interrupt table and a segment table.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Reference will now be made in detail to the preferred embodiment of the present invention, an example of which is illustrated in the accompanying drawings.

Fig. 1 is a schematic diagram of a network for an interactive simulated dialogue consistent with one embodiment of the present invention. In general, the network includes a client node 100 having a browser 110, an operating system 120, a client agent 130, and a client launch agent 140. The network further includes a server 160 and a server agent/launch agent 170. Client node 100 connects to

server 160 over a computer network 175 such as the Internet. Although the connection may be over any type of computer network, the computer network will hereinafter be referred to as the Internet for explanatory purposes.

Client node 100 is preferably an IBM-compatible personal computer with a Pentium-class processor, memory, and hard drive, preferably running Microsoft Windows. Generally, client node 100 also includes input and output components 102. Input components may include, for example, a mouse, keyboard, microphone, floppy disk drives, CD ROM and DVD drives. Output components may include, for example, a monitor, a sound card, and speakers. The monitor is preferably an XGA monitor with 1024 x 768 resolution and 16 bit color depth. The sound card may be a Sound Blaster or a comparable sound card. The number of client nodes is limited only by client license(s), available bandwidth, and hardware capability. For a detailed description of exemplary hardware components and implementation of client node 100, see U.S. Patent Nos. 5,006,987 and 5,730,603, to Harless.

Client agent 130 is a program that enables a user to ask a question in spoken, natural language and receive a meaningful response from a video character. The meaningful response is, for example, video and audio of the video character responding to the user's question. Client agent 130 preferably includes speech recognition software 180. Speech recognition software 180 is preferably one that is capable of processing a user's voice input. This eliminates the need to "train" the voice recognition software. An appropriate choice is Dragon Systems'

VoiceTools. Client agent 130 may also enable "intelligent prompting" as described below.

Operating system 120 connects to client launch agent 140 to oversee the checking and installation of necessary software and tools to enable client node 100 to run interactive simulated dialogues. While the process of checking and installing may be implemented at various stages, it is preferably performed for a first-time user during registration. Initially, a user at client node 100 may connect to server 160 via the Internet. The user then selects a case from a plurality of choices on server 160 through browser 110. Browser 110 sends the case-specific request to server launch agent 170. For first-time users, server launch agent 170 downloads and runs Csim Query 142 (explained in more detail in connection with Fig. 2).

Server 160 accesses database 162, which may be located at server 160 or a different location. Database 162 contains a vocabulary of questions or statements that may be understood by a virtual character in the selected case, and command words that allow the user to navigate through the program and review the session.

Database 162 also stores the plurality of interactive simulation scenarios. The interactive simulation scenarios are stored as a series of image frames on a media delivery device, preferably a CD ROM drive or a DVD drive. Each frame on the media delivery device is addressable and is accessible preferably in a maximum search time of 1.5 seconds. The video images may be compressed in a digital format, preferably using Intel's INDEO CODEC (compression/decompression

software) and stored on the media delivery device. Software located on the client node decompresses the video images for presentation so that no additional video boards are required beyond those in a standard multimedia configuration.

Database 162 preferably contains two groups of image frames. The first group relates to images of a story and characters involved in the simulated drama. The second group contains images providing a visual and textual knowledge base associated with the simulated topic, known as "intelligent prompts." Intelligent prompts may be used to also display scrolling questions, preferably three, that are dynamically selected for their relevance to the most recent response of the virtual character.

Server 160 further includes a server buffer agent, preferably video buffer agent 185 and scroll buffer agent 187. Client node 100 further includes a client buffer agent, preferably scroll buffer agent 191, video buffer agent 189, scroll pre-buffer 193, and video pre-buffer 195. These components are described in more detail below with reference to Fig 3.

Fig. 2 illustrates Csim Query 142. Csim Query 142 checks and installs the necessary software and tools to enable client node 100 to run interactive simulated dialogues. In step 210, server 160 interacts with client launch agent 140 using SPOT (SPeech On The web) 172 to determine whether a SAPI (Speech Applications Programmers Interface) compliant speech recognition engine, such as ViaVoice or Dragon Naturally Speaking™ resides on client node 100. SPOT 172 is a

commercial software program developed by Speech Solutions, Inc. If client node 100 does not have a SAPI compliant engine, client launch agent 140 determines if client node 100 has the minimum requirements to run the necessary software in step 212. If client node 100 has the minimum requirements to run the necessary software, client agent 140 downloads and installs the necessary software once permission is received in step 214. If client node 100 does not meet the minimum system requirements to run the software, the user is alerted and the install process is aborted in step 216.

If client launch agent 140 determines a SAPI compliant speech recognition engine resides on the system, client launch agent 140 then determines the identity and nature (version, level of performance, functionality) of the engine. If the engine has the recognition power (corpus size, independent speaker, continuous speech capabilities) and functionality (word spotting, vocabulary enhancement and customization), it is used by the interactive simulated dialogue program. If the resident engine does not have the recognition power and functionality to run the interactive simulated dialogue, client agent 140 downloads the necessary software once permission is received.

Once the necessary speech recognition software is installed on the user's system, client launch agent 140 determines if the case requested by the user is already on client node 100 as shown in step 218. If not, the files for the requested scenario are installed in step 220 on client node 100.

In step 222, client node 100 is optimized for user voice commands entered by, for example, a microphone. A Mic Volume Control Optimizer queries the client's operating system to determine its sound card specification, capabilities, and current volume control settings. Based on these finding, the optimizer adjusts the client system for voice commands. In a client node running Microsoft Windows, for example, the optimizer will create a backup of the current volume control settings in a temp directory and interface with the playback controls of the Windows volume control utility to deselect/mute the volume of the microphone playback through the client's speakers. The Mic Volume Control Optimizer also interfaces with a recording control of the Windows volume control utility to select and adjust the microphone input volume, and interfaces with the advanced controls of the microphone of the Windows volume control to enable the Mic gain input boost.

Fig. 3 is a general flow diagram of the interactive simulation consistent with one embodiment of the invention. A user, in step 305, selects a simulated dialogue program, or case. The user then connects to an Internet site and selects a simulated dialogue program by clicking with a mouse on an icon representing the desired program. As shown in step 307, the server than transmits to the client node a vocabulary set corresponding to the selected interactive simulation program.

The selected interactive simulation program allows the user to assume the role of, for example, a doctor diagnosing a patient. Using spoken inquires and

commands, the program allows the user to interview the patient/video character generated from images from database 162 and direct the course of action.

The simulated dialogue begins with an utterance or voice input by the user. As shown in step 310, the voice input is digitized and analyzed by the SAPI compliant speech recognition engine. The voice input may be prompted by comments, statements, or questions that scroll on the video display. The client agent, using the recognition engine (described in further detail below with reference to Fig. 4), then determines whether there is direct, indirect, or non-recognition of the utterance in step 320. Recognition of the voice input results in an interrupt number being sent by the client agent to the server agent (described in further detail with reference to Fig. 5). Server agent, in step 330, searches the internal database for a meaningful response for the video character. When a response is selected, its associated video segment consisting of image frames and audio signals representing human speech is retrieved and sent by the server video buffer agent to a client video buffer agent as shown in step 350. Prompts associated with the selected response are transmitted by the server scroll buffer agent to a client scroll buffer agent. In a preferred embodiment, three prompts are associated with a selected response. The prompts and video segments received by the client scroll and buffer agents are stored in a pre-buffer as shown in step 360. Using the monitor and speakers, client node 100 then plays the video and audio, and scrolls the prompts as shown in step 380. Upon seeing and hearing the meaningful

response to the user's question, the user continues the interactive simulated dialogue by entering another voice input based on the scrolling prompts.

In anticipation of the user's response of uttering another question based on the scrolling prompts, video segments and prompts associated with a meaningful response to the prompts are also downloaded from the server and buffered in the client system as shown in step 370. This minimizes response times to sustain the illusion of a continuous conversation with the character.

Fig. 4 illustrates the recognition engine of the client agent. A direct recognition 410 is almost always the result of the user selecting and uttering a phrase from the dynamic intelligent prompting system that scrolls the words and phrases from a precise vocabulary. These prompts help to guide a user unfamiliar with the system. If there is no direct recognition of the utterance, a second analysis ensues, using the logic and corpus of the resident recognition engine to determine what the user said. A second analysis is almost always required when the user utters a free speech inquiry that is either a paraphrase of a prompt or a spontaneous question or a statement that may or may not be answerable by the simulation character. In this second analysis, the text of the utterance is compared to a key word list of the instant scenario. If the comparison yields a match, the result is an indirect recognition 420. If the comparison does not yield a match 430, the text of the utterance is transmitted through the Internet interface to the server agent with a parameter indicating that the utterance could not be understood 440. A direct or

indirect recognition results in an interrupt number being sent through the Internet interface to server agent 330 explained in further detail with respect to Fig. 5.

In order to avoid displaying redundant prompts that will trigger redundant scenes, interrupt handler 450 maintains a list of previously displayed scene segments. In the event an utterance is mis-recognized as redundant, mis-recognition segment buffer 460 buffers video segments that inform the user that an utterance was not recognized.

Fig. 5 illustrates in further detail the step of receiving an interrupt number by the server agent (step 330 of Fig. 3). Reception of an interrupt number by interrupt agent 510 initiates a search of database 562 for a meaningful response from the video character. When a response is selected, the response and its associated prompts are transmitted to scroll buffer agent 587. The associated video segment are also retrieved and transmitted it to the video buffer agent 585. As previously discussed, video buffer agent 587 also retrieves video segments associated with subsequent responses to the transmitted prompts. In one embodiment, video buffer agent 587 determines the network capacity for the transfer of the video segments. Network capacity depends on many factors including available bandwidth and network connection speed. Based on this determination, video agent 587 transfers portions of the video segments of each of the subsequent responses on a rotational basis. Since video agent 587 rotates only the relevant segments to the most recent response into the buffer, download time is minimized and bandwidth saved.

Fig.6 illustrates in further detail the step of selecting an interrupt number in response to the user's utterance (step 330 of Fig. 3). In each interactive simulation, a potential topic of conversation is assigned a state 610. There is no limit to the number of states that can exist for a given interactive simulation. State 610, for example, relates to medical history. Within each state are suggested questions 620 that prompt the user to elicit a response from the video character. If a user utters a prompted phrase that is recognized by the recognition engine, an interrupt number is transmitted to the interrupt agent. Interrupt table 630, as shown in Fig. 6, contains segment numbers 635 which point to corresponding segment numbers 645 in a segment table 640. For example, the first segment number "0006" of interrupt table 630 points to segment number "0006" of segment table 640. Each segment number 645 of segment table 640 corresponds to a particular scene stored on the media delivery device. The video agent at the direction of the interrupt agent retrieves the video segment corresponding to the referenced segment and outputs it to the video buffer.

Referring again to Fig. 1, the processor of client node 100 executes one or more sequences of one or more instructions contained in the memory. Such instructions may be read into the memory from a computer-readable medium via input/output device 102. Execution of the sequences of instructions contained in the memory causes the processor to perform the process steps described herein. In an alternative implementation, hard-wired circuitry may be used in place of or in

combination with software instructions to implement the invention. Thus implementations of the invention are not limited to any specific combination of hardware circuitry and software.

The term "computer-readable medium" as used herein refers to any media that participates in providing instructions to the processor of client node 100 for execution. Such a medium may take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media includes, for example, optical or magnetic disks. Volatile media includes dynamic memory. Transmission media includes coaxial cables, copper wire, and fiber optics. Transmission media can also take the form of acoustic or light waves, such as those generated during radio-wave and infra-red data communications.

Common forms of computer-readable media include, for example, a floppy disk, a flexible disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, punch cards, papertape, any other physical medium with patterns of holes, a RAM, PROM, and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave as described hereinafter, or any other medium from which a computer can read. Network signals carrying digital data, and possibly program code, to and from client node 100 are exemplary forms of carrier waves transporting the information. In accordance with the present invention, program code received by client node 100 may be executed by the

processor as it is received, and/or stored in memory, or other non-volatile storage for later execution.

It will be apparent to those skilled in the art that various modifications and variations can be made in the interactive audiovisual simulation system and method of the present invention and in construction of this system without departing from the scope or spirit of the invention.

Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.